HOW GOOD IS YOUR SUPER-RESOLUTION IMAGE? QUALITY ASSURANCE IN IMAGE RECONSTRUCTION USING THE BOOTSTRAP

Christian Debes

AGT International
64295 Darmstadt, Germany
E-mail: cdebes@agtinternational.com

Christian Weiss, Abdelhak M. Zoubir

Signal Processing Group
Technische Universität Darmstadt
64283 Darmstadt, Germany
E-mail: {cweiss, zoubir}@spg.tu-darmstadt.de

ABSTRACT

Super-Resolution Image Reconstruction is known to be sensitive to errors in assumptions such as accurate sub-pixel motion estimation. Even small errors can yield a significant degradation of image quality that complicates any follow-on task such as object detection or classification. We focus on the problem of automatic quality assessment of Super-Resolution image reconstruction. We propose a bootstrap-based method that provides an objective metric quantifying reconstruction quality and thus allowing to readjust the reconstruction.

1. INTRODUCTION

Super-Resolution Image Reconstruction (SRIR) techniques [1] allow us to obtain a high-resolution (HR) image from a set of spatially displaced low-resolution (LR) images. SRIR is thus of high importance in areas where sensing a set of images with reduced resolution is more practical, cheaper, less dangerous or faster than acquiring a single image of high resolution. In a nutshell, SRIR replaces expensive sensing hardware by computational power. This is of interest in applications such as remote sensing, surveillance and medical imaging where it has successfully been applied in the past.

In order to successfully 'fuse' and interpolate the set of LR images to a single HR image, estimates of the projections from the HR image to each of the LR images are crucial. Depending on the practical scenario, these projections include, among others, motion, blurring, warping, lighting changes and zooming. Further, noise models are necessary to e.g. perform maximum likelihood (ML) or maximum a posteriori (MAP) estimation [1, 2, 3]. It is noted that some estimates have a higher impact on the success of SRIR than others. High-precision subpixel motion estimation for example is crucial to correctly align the set of LR images. Even small errors in the motion estimate can show dramatic effects on the resulting HR image. Errors in estimating the lighting change on the other hand will generally not show significant impact on the image quality.

For a practical SRIR system, automatic reconstruction is a key issue. That is, given a set of LR images, the corresponding projections are to be estimated from the data at hand, followed by the actual image reconstruction step. Automatic quality assessment of the image reconstruction is then of crucial importance. Assigning quality information to image reconstruction results can be of practical help to e.g.

- re-estimate the projection matrices using more advanced methods (for example flow-based instead of simple block-based motion estimation)
- use Super-Resolution techniques only for those image parts where, due to sufficient simplicity of the projection matrices, it can successfully be applied, and use classical interpolation for the rest of the image
- use only a subset of the LR images for reconstruction
- decide not to use Super-Resolution at all.

There is little work done in assessing the quality of SRIR. Mostly, researchers deal with quality assessment using reference images to determine e.g. the most suitable image reconstruction algorithm [4, 5, 6]. Further, there exist a variety of literature in the area of robust SRIR that allows to perform image reconstruction when deviations from the assumed projections occur [7, 8]. The only paper, to the best of our knowledge, that at least partly deals with the situation when no reference image is available is [9], wherein the authors use a metric to determine the degree of ringing and blurring in the SRIR images. This metric is, however, highly image- and contamination-dependent as it implicitly assumes that blurring and high frequencies occur when SRIR is not successful. However, both effects can naturally occur in images which makes the method in [9] of limited use in practice. Only the reference image based metrics presented in [9], such as the mean square error (MSE) and the structural similarity index measure (SSIM) [10] have been shown to reliably match the obtained image quality.

The contribution of this paper is the introduction of a fully automatic framework for assessing the quality of Super-Resolution Image Reconstruction without assuming a reference image to be available. We make use of the bootstrap principle [11, 12] to resample with replacement from image reconstruction residuals to estimate characteristics of the probability density function (pdf) of the super-resolved image. Given an estimate of such property, we demonstrate how quality information of the conducted image reconstruction can be acquired.

2. SIGNAL MODEL

Let the set of acquired LR images be denoted by $\{\mathbf{y}_l\}_{l=0}^{L-1}$ where L denotes the total number of acquired images. Here, every $\mathbf{y}_l$ with $l = 0, ..., L-1$ is a $(N \cdot M) \times 1$ vector that represents the vectorized version of the two-dimensional $N \times M$ image in lexicographic notation. The linear signal model for each $\mathbf{y}_l$, $l = 0, ..., L-1$ is then given by,

$$\mathbf{y}_l = \mathbf{A}_l \mathbf{x} + \mathbf{n}_l \tag{1}$$

where $\mathbf{x}$ denotes the true underlying HR image of dimension $(p^2 \cdot N \cdot M) \times 1$ with p being the downsampling factor. The $(N \cdot M) \times (p^2 \cdot N \cdot M)$ matrix $\mathbf{A}_l$ performs the projection from the high-dimensional $\mathbf{x}$ to the low-dimensional $\mathbf{y}_l$ for $l = 0, ..., L-1$ and the $(N \cdot M) \times 1$ vector $\mathbf{n}_l$ is the additive noise component.

The projection matrix $\mathbf{A}_l$ is a product of several matrices that treat motion, blurring, downsampling, and other distortions individually. A common modelling [1] is

$$\mathbf{A}_l = \mathbf{D} \cdot \mathbf{B} \cdot \mathbf{M}_l \tag{2}$$

where $\mathbf{D}$ is the $(N \cdot M) \times (p^2 \cdot N \cdot M)$ matrix that performs the downsampling by p in both image dimensions. The blurring matrix of size $(p^2 \cdot N \cdot M) \times (p^2 \cdot N \cdot M)$ is denoted as $\mathbf{B}$ and is related to the Point Spread Function (PSF) of the imaging system. Finally, $\mathbf{M}_l$ is the $(p^2 \cdot N \cdot M) \times (p^2 \cdot N \cdot M)$ motion matrix that describes the relative motion between $\mathbf{x}$ and the accordingly interpolated $\mathbf{y}_l$. The motion matrix operates in the high-dimensional space which relates to sub-pixel displacements of the LR images. Those sub-pixel displacements and the hereby introduced aliasing effects are of fundamental importance in SRIR. Note that in Equation (2) the downsampling and blurring matrices are independent of the image index l, $l = 0, ..., L-1$.

3. SUPER-RESOLUTION IMAGE RECONSTRUCTION

A large variety of SRIR algorithms exists, including the intuitive non-uniform interpolation [13], frequency-domain approaches [14] and stochastic regularization [3], to name a few. In this paper, we focus on stochastic regularization approaches. It is noted, however, that the proposed quality assessment framework to be presented in Section 4 is independent of the actual reconstruction technique and can straightforwardly be applied to any SRIR technique.

Consider the linear signal model in Equation (1). Under the assumption of $\mathbf{n}_l$ being i.i.d. Gaussian noise that is also independent with the image index l, the Maximum *A Posteriori* (MAP) estimator can be written as,

$$\hat{\mathbf{x}} = \arg\max_{\mathbf{x}} p(\mathbf{x}|\mathbf{y}_0, ..., \mathbf{y}_{L-1}) \tag{3}$$

leading to the following optimization problem [3]

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \left\{ \sum_{l=0}^{L-1} ||\mathbf{y}_l - \mathbf{A}_l \mathbf{x}||^2 + \rho \cdot \phi(\mathbf{x}) \right\} \tag{4}$$

where ρ is the regularization parameter and $\phi(\mathbf{x})$ includes the Markov-Random-Field (MRF) - prior that is typically implemented as a function of the image derivative [15]. Equation (4) can efficiently be implemented by gradient descent algorithms such as the conjugate gradient (CG) algorithm [16].

It is noted that in practice, the matrices $\mathbf{A}_l$, $l = 0, ..., L-1$ in Equation (4) are replaced by their estimates $\hat{\mathbf{A}}_l$, $l = 0, ..., L-1$. The estimates are obtained by independently or jointly estimating e.g. the blurring and most importantly the motion between the LR images [17, 18, 19].

4. QUALITY ASSESSMENT

It is known [1] that the success of SRIR is highly dependent on the estimation quality of the projection matrices $\mathbf{A}_l$, $l = 0, ..., L-1$. In scenarios where accurate estimates of these matrices cannot be obtained, the image quality drastically degrades, introducing strong artifacts that deteriorate any follow-on task such as detection, classification or segmentation.

The output of SRIR, e.g. by use of Equation (4) is an estimate of the true HR image, $\mathbf{x}$. A meaningful objective quality metric should be a function of this estimate $\hat{\mathbf{x}}$. We note that $\hat{\mathbf{x}}$ is a realization of a random vector $\hat{\mathbf{X}}$ that follows an (unknown) pdf $f_{\hat{\mathbf{X}}}(\hat{\mathbf{x}})$. This pdf displays the variability of the image reconstruction, given the observations $\{\mathbf{y}_l\}_{l=0}^{L-1}$ which again are drawn from the (unknown) multivariate distribution $f_{\mathbf{Y}_1, \mathbf{Y}_2, ..., \mathbf{Y}_L}(\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_L)$.

The pdf of $\hat{\mathbf{X}}$ gives strong insight into the quality of SRIR. The ideal pdf of the HR image estimate woud be $f_{\hat{\mathbf{X}}}(\hat{\mathbf{x}}) = \delta(\hat{\mathbf{x}} - \mathbf{x})$ where δ is Dirac's delta. This would allow for perfect image reconstruction. The more the actual pdf deviates from a highly concentrated pdf, the lower is the confidence in the quality of SRIR. We can thus re-formulate the problem of quality assessment of SRIR by estimating a corresponding characteristic of the pdf of the HR image estimate $f_{\hat{\mathbf{X}}}(\hat{\mathbf{x}})$ and evaluating its variability or any other measure for the deviation from a Dirac delta.

Ideally, the pdf $f_{\hat{\mathbf{X}}}(\hat{\mathbf{x}})$ would be obtained by running the experiment independently many times, estimating $\mathbf{x}$ for each of those experiments and based on the set of estimated HR images estimate the density $f_{\hat{\mathbf{X}}}(\hat{\mathbf{x}})$. 'Running the experiment many times' in the SRIR scenario would mean to fix the set of projection matrices $\{\mathbf{A}_l\}_{l=0}^{L-1}$ as well as the noise density and then draw i.i.d. LR images $\{\mathbf{y}_l\}_{l=0}^{L-1}$. Clearly, this approach is infeasible in practice as the experimental conditions are typically non-stationary and keeping the projection matrices $\mathbf{A}_l$ constant for all l is highly impractical. In most SRIR applications, one is faced with a single set of measurements only, from which inference shall be drawn.

The bootstrap [20, 11, 12] is a tool that allows to draw statistical inference by resampling with replacement from few measurements. It is especially useful in applications where few data samples are available, experiments cannot be repeated and no analytical solutions are available. The bootstrap has been applied successfully to estimate, e.g., confidence intervals of detection results and parameter estimates, e.g. in radar imaging [21] and objective speech quality enhancement [22]. In the sequel, we describe a bootstrap based method to estimate the objective characteristic of the pdf $f_{\hat{\mathbf{X}}}(\hat{\mathbf{x}})$ from a single set of measurements $\{\mathbf{y}_l\}_{l=0}^{L-1}$ as well as, based on it, draw conclusions on the relative quality of different low-resolution data sets used for SRIR. Consider the linear signal model in Equation (1). Given an estimate of the HR image $\mathbf{x}$ - obtained by any of the existing SRIR techniques - one can obtain an estimate of the noise vectors $\mathbf{n}_l$, $l = 0, ..., L-1$ as

$$\hat{\mathbf{n}}_l = \mathbf{y}_l - \mathbf{A}_l\hat{\mathbf{x}} \tag{5}$$

for $l = 0, ..., L-1$. As $\mathbf{n}_l$ per definition is i.i.d. for all l, we propose resampling with replacement from the residuals $\hat{\mathbf{n}}_l$ for all l and thus obtaining pseudo noise data $\hat{\mathbf{n}}_l^b$ with $l = 0, ..., L-1$ and $b = 0, ..., B-1$ where B denotes the total number of bootstrap resamples. The pseudo noise data can in turn be considered to generate pseudo LR images as

$$\mathbf{y}_l^b = \mathbf{A}_l\hat{\mathbf{x}} + \hat{\mathbf{n}}_l^b \tag{6}$$

resulting in B new sets of LR images $\{\mathbf{y}_l^b\}_{l=0}^{L-1}$ with $b = 0, ..., B-1$. These B sets can now be used independently for SRIR to obtain B pseudo HR images $\hat{\mathbf{x}}^b$ with $b = 0, ..., B-1$, from which the desired characteristic of $f_{\hat{\mathbf{X}}}(\hat{\mathbf{x}})$ can be estimated. It is known that for $L \to \infty$ the pseudo data pdf approaches the true pdf $f_{\hat{\mathbf{X}}}(\hat{\mathbf{x}})$, given the signal model holds. The variability of $f_{\hat{\mathbf{X}}^{\mathbf{b}}}(\hat{\mathbf{x}}^{\mathbf{b}})$ when different sets of low-resolution images are used, can now be used as a measure of relative image reconstruction quality, i.e. the more $f_{\hat{\mathbf{X}}^{\mathbf{b}}}(\hat{\mathbf{x}}^{\mathbf{b}})$ deviates from Dirac's delta the less reliable is the reconstructed image $\hat{\mathbf{x}}$. A simple metric that describes this characteristic and which can be used to infer relative quality information is the mean intensity variance (MIV)

$$\text{MIV}(\hat{\mathbf{x}}) = \frac{1}{NM}\sum_{n=0}^{NM}\frac{1}{B}\left(\hat{\mathbf{x}}^b(n) - \frac{1}{B}\sum_{b=0}^{B-1}\hat{\mathbf{x}}^b(n)\right)^2 \tag{7}$$

which displays the mean variance of the pixel intensity over all bootstrap realizations.

Table 1. The Bootstrap procedure

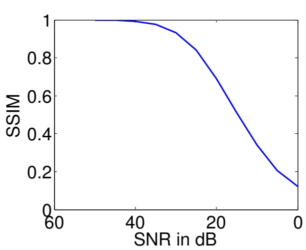
Step 0.	*Data Collection*. Conduct the experiment and collect L LR images $\mathbf{y}_l$, $l = 0, ..., L-1$.
Step 1.	*Preprocessing*. Estimate the projection matrices $\mathbf{A}_l$, $l = 0, ..., L-1$ using e.g. motion and blur estimation techniques.
Step 2.	*Image Reconstruction*. Estimate an HR image $\hat{\mathbf{x}}$, e.g. by considering a MAP approach as $$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}}\left\{\sum_{l=0}^{L-1}\|\mathbf{y}_l - \mathbf{A}_l\mathbf{x}\|^2 + \rho\cdot\phi(\mathbf{x})\right\}$$
Step 3.	*Resampling*. Calculate $\hat{\mathbf{n}}_l = \mathbf{y}_l - \mathbf{A}_l\hat{\mathbf{x}}$, $l = 0, ..., L-1$ and generate B bootstrap realizations $\hat{\mathbf{n}}_l^b$, $l = 0, ..., L-1$ and $b = 0, ..., B-1$ by resampling with replacement from $\hat{\mathbf{n}}_l$, $l = 0, ..., L-1$.
Step 4.	*Generate pseudo LR images*. Calculate B sets of pseudo LR images as $\mathbf{y}_l^b = \mathbf{A}_l\hat{\mathbf{x}} + \hat{\mathbf{n}}_l^b$, $l = 0, ..., L-1$ and $b = 0, ..., B-1$.
Step 5.	*Bootstrap HR images*. Calculate B pseudo HR images as $$\hat{\mathbf{x}}^b = \arg\min_{\mathbf{x}}\left\{\sum_{l=0}^{L-1}\|\mathbf{y}_l^b - \mathbf{A}_l\mathbf{x}\|^2 + \rho\cdot\phi(\mathbf{x})\right\}$$ with $b = 0, ..., B-1$
Step 6.	*Quality metric estimation*. Estimate the mean intensity variance as $$\text{MIV}(\hat{\mathbf{x}}) = \frac{1}{NM}\sum_{n=0}^{NM}\frac{1}{B}\left(\hat{\mathbf{x}}^b(n) - \frac{1}{B}\sum_{b=0}^{B-1}\hat{\mathbf{x}}^b(n)\right)^2$$

The bootstrap procedure is summarized in Table 1 whereby we restricted ourselves to the MAP approach for SRIR and the MIV as quality metric, noting, however, that the actual procedure of estimating the proposed characteristic of $f_{\hat{\mathbf{X}}}(\hat{\mathbf{x}})$ is independent of the considered SRIR technique and also different quality metrics can easily be extracted from the density $f_{\hat{\mathbf{X}}^{\mathbf{b}}}(\hat{\mathbf{x}}^{\mathbf{b}})$.

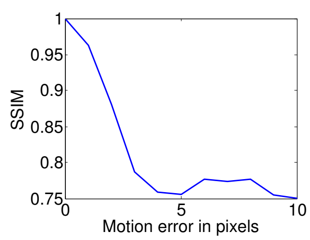
5. RESULTS

We consider a set of different simulation setups in order to validate the performance of the proposed quality metric and its ability to map the degree of contamination in LR images. In the first simulation setup, we consider two LR images obtained from one HR image by a downsampling factor of $p = 2$, Gaussian blur and a relative shift of 0.5 pixels.

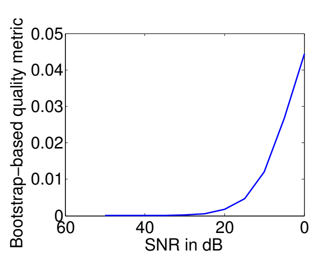
To show the performance of the proposed bootstrap-based quality metric, we choose the well-known SSIM [10] for comparision. Note that the SSIM is a reference-based image quality metric that needs a successful image reconstruction to be available. This is never the case in practice and as such, the SSIM is considered as clairvoyant metric that can only be used in a fully controlled simulation setup. The bootstrap-based quality metric, on the other hand, is not based on the assumption of having a reference image available but draws inference from the data at hand. Such a direct comparison of the two approaches is meaningless in terms of absolute image quality, but the performance of the proposed approach can be measured in terms of how well our results of inappropriate combinations of low-resolution images match with the results from a reference image-based approach.



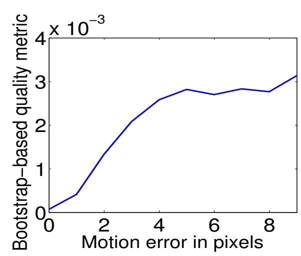
(a) SSIM versus noise



(b) SSIM vs. motion errors



(c) Proposed metric versus noise



(d) Proposed metric versus motion errors

Fig. 1. Comparison of the proposed bootstrap-based quality metric and the SSIM

We choose additive white Gaussian noise and translational motion errors as contaminations to be examined in the following. In Figure 1, the two metrics are plotted as a function of the SNR using 100 Monte Carlo simulations and the translational motion. Considering the additive noise, the SSIM in Figure 1(a) remains close to unity for high SNR and then linearly decreases with decreasing SNR. The proposed metric in Figure 1(c) is consistent with this observation. Firstly, it is noted that the bootstrap-based metric monotonically increases with increasing contamination. Secondly, it shows a similar trend as the reference image-based SSIM. It is approximately constant for high SNR and then linearly increases with decreasing SNR.

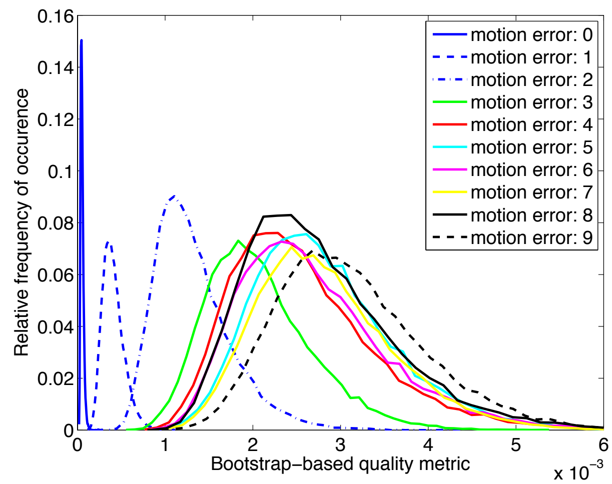


Fig. 2. Histogram of the bootstrap-based quality metric intensity variances for data sets with different levels of motion errors

A similar effect can be observed when considering motion errors in Figures 1(b) and 1(d). As expected, the SSIM strongly decreases when adding few motion errors and then remains approximately at a constant (poor) quality. The bootstrap-based quality metric in Figure 1(d) on the other hand quickly increases for few motion errors and then also remains approximately constant.

Figure 2 depicts the histogram of the pixel intensity variances in HR images estimated from B sets of bootstrap-based low-resolution images. We consider different motion errors for one low-resolution image in each data set. Again, we can observe that the bootstrap-based quality metric correctly maps the underlying degree of contamination.

6. CONCLUSION

An objective metric for quality estimation of Super-Resolution image reconstruction has been presented. It is based on the idea of resampling image residuals using the bootstrap to obtain pseudo high resolution images. By doing so, a probablitiy density function of the high resolution images is obtained, from which a quality metric is infered. The proposed techniques are independent of the actual reconstruction algorithm and allow for an automatic framework for image reconstruction with quality assurance. Results on simulated data have shown that the bootstrap-based quality metric shows a strong correlation with reference-based metrics such as SSIM.

7. REFERENCES

[1] S.C. Park, M.K. Park, and M.G. Kang, "Super-resolution image reconstruction: a technical overview," *IEEE Signal Processing Magazine*, vol. 20, no. 3, pp. 21–36, 2003.

[2] T.L.M. Kasteren, G. Englebienne, and B. J.A. Kröse, "Hierarchical activity recognition using automatically clustered actions," in *Ambient Intelligence*, D. V. et al Keyson, Ed., vol. 7040 of *Lecture Notes in Computer Science*, pp. 82–91. Springer Berlin Heidelberg, 2011.

[3] R. R. Schultz and R. L. Stevenson, "Extraction of high-resolution frames from video sequences," *IEEE Transactions on Image Processing*, vol. 5, no. 6, pp. 996–1011, 1996.

[4] T. Ahmad and S. S. Quershi, "The full reference quality assessment metrics for super resolution of an image: Shedding light or casting shadows?," in *Proc. Int Electronics and Information Engineering (ICEIE) Conf. On*, 2010, vol. 2.

[5] I. Begin and F. P. Ferrie, "Comparison of super-resolution algorithms using image quality measures," in *Proc. 3rd Canadian Conf. Computer and Robot Vision*, 2006.

[6] X. Zhou and B. Bhanu, "Evaluating the quality of super-resolved images for face recognition," in *Proc. IEEE Computer Society Conf. Computer Vision and Pattern Recognition Workshops CVPRW '08*, 2008, pp. 1–8.

[7] M. K. Ng and N. K. Bose, "Analysis of displacement errors in high-resolution image reconstruction with multisensors," *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, vol. 49, no. 6, pp. 806–813, 2002.

[8] H. Yan, J. Liu, J. Sun, and X. Ji, "Regularization super-resolution image fusion considering inaccurate image registration and observation noise," in *Proc. Int Neural Networks and Signal Processing Conf*, 2008, pp. 91–94.

[9] A. R. Reibman, R. M. Bell, and S. Gray, "Quality assessment for super-resolution image enhancement," in *Proc. IEEE Int Image Processing Conf*, 2006, pp. 2017–2020.

[10] Z. Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, April 2004.

[11] A.M. Zoubir and D. R. Iskander, *Bootstrap Techniques for Signal Processing*, Cambridge University Press, 2004.

[12] A.M. Zoubir and D.R. Iskander, "Bootstrap methods and applications," *IEEE Signal Processing Magazine*, vol. 24, no. 4, pp. 10–19, 2007.

[13] J. Clark, M. Palmer, and P. Lawrence, "A transformation method for the reconstruction of functions from nonuniformly spaced samples," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 33, no. 5, pp. 1151–1165, 1985.

[14] S. P. Kim, N. K. Bose, and H. M. Valenzuela, "Recursive reconstruction of high resolution image from noisy undersampled multiframes," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 38, no. 6, pp. 1013–1027, 1990.

[15] G. Winkler, *Image Analysis, Random Fields and Markov Chain Monte Carlo Methods*, Springer, 2003.

[16] R. Fletcher, "Function minimization by conjugate gradients," *Computer Journal*, vol. 7, no. 2, pp. 149–154, 1964.

[17] B. Furht, J. Greenberg, and R. Westwater, *Motion Estimation algorithms for Video Compression*, Kluwer Academic Publishers, 1997.

[18] R. C. Hardie, K. J. Barnard, and E. E. Armstrong, "Joint MAP registration and high-resolution image estimation using a sequence of undersampled images," *IEEE Transactions on Image Processing*, vol. 6, no. 12, pp. 1621–1633, 1997.

[19] C. Debes, T. Wedi, C. L. Brown, and A. M. Zoubir, "Motion estimation using a joint optimisation of the motion vector field and a super-resolution reference image," in *Proc. IEEE Int. Conf. Image Processing ICIP 2007*, 2007, vol. 2.

[20] B. Efron and R.J. Tibshirani, *An Introduction to the Bootstrap*, Chapman & Hall, 1993.

[21] C. Debes, C. Weiss, A. M. Zoubir, and M. G. Amin, "Distributed target detection in through-the-wall radar imaging using the bootstrap," in *Proc. IEEE Int Acoustics Speech and Signal Processing (ICASSP) Conf*, 2010, pp. 3530–3533.

[22] Zhihua Lu, P. Heidenreich, and A. M. Zoubir, "Objective quality assessment of speech enhancement algorithms using bootstrap-based multiple hypotheses tests," in *Proc. IEEE Int Acoustics Speech and Signal Processing (ICASSP) Conf*, 2010, pp. 4102–4105.